

# Nucleotide sequence of a mouse cDNA encoding the non-histone chromosomal high mobility group protein-1 (HMG1)

Wagner V.Yotov and René St-Arnaud

Division of Surgical Research, McGill University and Genetics Unit, Shriners Hospital for Crippled Children, Montreal, Quebec H3G 1A6, Canada

Submitted May 21, 1992

EMBL accession no. Z11997

The high mobility group (HMG) proteins are the major non-histone components of chromatin (1). Among these, the HMG1 and HMG2 proteins are the most abundant; however, their precise function remains undefined. There are some indications for a putative role of HMG1/2 in chromatin assembly (2), DNA replication (3) or even as general transcription factors for RNA polymerase II (4).

There is a considerable body of recent literature showing similarity between particular transcription factors and HMG1/2. The homologous regions of these proteins are known as HMG boxes. Among these factors are hUBF (human upstream binding factor) (5), SRY (mammalian testes-determining factor) (6), LEF-1 (lymphoid enhancer-binding factor 1) (7), TCF-1 (T cell-specific factor) (8) and their analogues in other species.

The progress in the study of HMG box-containing proteins is in contrast with the unavailability of HMG1 sequencing data for such a widely used laboratory model as the mouse. There is also confusion in the available sequences, caused by the significant differences in the length of the 5' and 3' non-coding regions. We addressed this problem by cloning and sequencing mouse HMG1 cDNAs. We constructed a cDNA library into the lambda-Zap vector (Stratagene Corp., La Jolla, CA) using mRNA isolated from the P19 line of embryonal carcinoma cells (9). The sequence was obtained from three distinct overlapping clones (1-793, 1-1229, 1176-2231) and both strands of the recombinant cDNAs were sequenced entirely. This is the longest available HMG1 cDNA sequence. The ORF (nucleotides 73-718) encodes a 215 amino acid-long protein.

When compared to the published cDNA sequences of other species, the mouse sequence shows the strongest similarity with HMG1 cDNAs from other rodents (see table 1). Mammalian HMG1 proteins contain two HMG boxes (amino acid positions 9-90 and 95-173). The only differences at the protein level are localised outside of these boxes. On the contrary, non-mammalian HMG protein sequences are less conserved, as are the HMG box-containing transcription factors. This may be due to the fact that HMGs and HMG box-containing proteins share only the necessity to bind DNA, but differ significantly in the role and specificity of this binding.

Our HMG1 recombinants could prove helpful for the expression of recombinant mouse HMG1 protein and for further studies of its still unclear function.

## ACKNOWLEDGEMENTS

We thank E.Wan for the synthesis of the oligonucleotides used for DNA sequencing. This project was supported by a grant from

the Shriners of North America (project no. 15964). R.St-A. is a chercheur-boursier from the Fonds de la Recherche en Santé du Québec.

## REFERENCES

1. Johns, E.W. (1982) In: *The HMG chromosomal proteins*, Academic Press, N.Y. p. 251.
2. Kohlsaat, L.A., et al. (1986) *J. Biol. Chem.* 261, 524-526.
3. Bonne-Andrea, C., et al. (1984) *EMBO J.* 5, 1193-1199.
4. Singh, J. and Dixon, G.H. (1990) *Biochemistry* 29, 6295-6302.
5. Jantzen, H.-M., et al. (1990) *Nature* 344, 830-836.
6. Sinclair, A.H., et al. (1990) *Nature* 346, 240-244.
7. Travis, A., et al. (1991) *Genes Dev.* 5, 880-894.
8. van de Wetering, M., et al. (1991) *EMBO J.* 10, 123-132.
9. McBurney, M.W., et al. (1982) *Nature* 299, 165-167.

Table 1.

Species	Gene/ Protein	Length		Similarity			
		nucl.	aa	nucleotides coding region	3'-end	amino acids identity	amino acids similarity
<i>Mammalian HMG1 genes and proteins</i>							
Mouse	HMG1	2231	215	100	100	100	100
Rat	HMG1	825	215	97.7	99.3	100	100
C.hamster	HMG1	2028	164	95.2	83.7	100	100
Human	HMG1	1009	215	87.2	98.6	99.1	100
Bovine	HMG1	1236	215	88.31	82.4	98.6	100
Pig	HMG1	2192	215	86.2	79.9	98.1	99.5
<i>Other HMG proteins</i>							
Trout	HMG1		204			89.5	95.1
Drosophila	HMG1		112			42.3	60.0
Trypanosoma	HMG1		271			28.2	50.0
Maize	PMR		157			41.0	47.4
Soybean	HMG1		152			37.2	46.1
T.thermophila	HMG1		64			19.2	37.2
S.cerevisiae	ACP2		251			20.5	34.6
Human	HMG17		100			20.5	33.3
Human	HMG14		90			15.4	33.3
Human	HMG1		156			14.6	29.3
<i>Other HMG box-containing proteins (see text)</i>							
Human	hUBP		764			24.3	45.2
Mouse	TCF-1		303			21.8	42.3
Mouse	LEF-1		397			21.8	41.0

Nucleotide and amino acid sequence comparison between the mouse HMG1 sequence and different HMG-related genes and/or proteins. For proteins other than HMG1, the amino acid comparisons were made only between their HMG-box(es) and the two HMG boxes of the mouse HMG1 protein. The result with higher similarity is shown in the columns. MacMol software (Soft Gene GmbH, Berlin, Germany) was used as a sequencing data source and an alignment tool.